

Assignment 3: Data Exploration

Summer Heschong

Fall 2024

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change “Student Name” on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Assign a useful **name to each code chunk** and include ample **comments** with your code.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
7. After Knitting, submit the completed exercise (PDF file) to the dropbox in Canvas.

TIP: If your code extends past the page when knit, tidy your code by manually inserting line breaks.

TIP: If your code fails to knit, check that no `install.packages()` or `View()` commands exist in your code.

Set up your R session

1. Load necessary packages (tidyverse, lubridate, here), check your current working directory and upload two datasets: the ECOTOX neonicotinoid dataset (ECOTOX_Neonicotinoids_Insects_raw.csv) and the Niwot Ridge NEON dataset for litter and woody debris (NEON_NIWO_Litter_massdata_2018-08_raw.csv). Name these datasets “Neonics” and “Litter”, respectively. Be sure to include the sub-command to read strings in as factors.

```
#setting up
library(tidyverse); library(lubridate); library(here) #load packages

here() #check working directory
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```
#load data sets:
Neonics <- read.csv(
  file = here("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv"),
  stringsAsFactors = TRUE)  #load and name neonics

#view(Neonics)

Litter <- read.csv(
  file = here("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv"),
  stringsAsFactors = TRUE)  #load and name litter

#view(Litter)
```

Learn about your system

2. The neonicotinoid dataset was collected from the Environmental Protection Agency's ECOTOX Knowledgebase, a database for ecotoxicology research. Neonicotinoids are a class of insecticides used widely in agriculture. The dataset that has been pulled includes all studies published on insects. Why might we be interested in the ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

Answer: We might be interested in this topic if we were researching certain insects or the effect of neonicotinoids on ecosystems more generally. These chemicals are harmful to many beneficial insects like pollinators and aquatic invertebrates, both of which are integral to the health of their ecosystems.

3. The Niwot Ridge litter and woody debris dataset was collected from the National Ecological Observatory Network, which collectively includes 81 aquatic and terrestrial sites across 20 ecoclimatic domains. 32 of these sites sample forest litter and woody debris, and we will focus on the Niwot Ridge long-term ecological research (LTER) station in Colorado. Why might we be interested in studying litter and woody debris that falls to the ground in forests? Feel free to do a brief internet search if you feel you need more background information.

Answer: So I actually was working at NEON before coming to school! I was on the flora team so I did both the litter and fine woody debris protocols. I'm so excited that we're using their data! Litter is very important in nutrient cycling in a forest; so we might be interested to know how much carbon for example is being mineralized through the litter.

4. How is litter and woody debris sampled as part of the NEON network? Read the [NEON_Litterfall_UserGuide.pdf](#) document to learn more. List three pieces of salient information about the sampling methods here:

Answer: 1. Elevated traps are set out that catch litter. The litter is collected from them in the spring and fall 2. Ground traps are designated rectangles of ground where fine woody debris is collected once a year 3. The samples of litter are taken back to the lab and sorted by type i.e. leaves, seeds, flowers, etc.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#dimensions of dataset?
colnames(Neonics) #column names of dataset
```

```
## [1] "CAS.Number" "Chemical.Name"
## [3] "Chemical.Grade" "Chemical.Analysis.Method"
## [5] "Chemical.Purity" "Species.Scientific.Name"
## [7] "Species.Common.Name" "Species.Group"
## [9] "Organism.Lifestage" "Organism.Age"
## [11] "Organism.Age.Units" "Exposure.Type"
## [13] "Media.Type" "Test.Location"
## [15] "Number.of.Doses" "Conc.1.Type..Author."
## [17] "Conc.1..Author." "Conc.1.Units..Author."
## [19] "Effect" "Effect.Measurement"
## [21] "Endpoint" "Response.Site"
## [23] "Observed.Duration..Days." "Observed.Duration.Units..Days."
## [25] "Author" "Reference.Number"
## [27] "Title" "Source"
## [29] "Publication.Year" "Summary.of.Additional.Parameters"
```

```
str(Neonics) #structure and dimensions of dataset
```

```
## 'data.frame': 4623 obs. of 30 variables:
## $ CAS.Number : int 58842209 58842209 58842209 58842209 58842209 58842209 58842209 58842209
## $ Chemical.Name : Factor w/ 9 levels "(1E)-N-[(6-Chloro-3-pyridinyl)methyl]-N-ethy"
## $ Chemical.Grade : Factor w/ 9 levels "Analytical grade",...: 9 9 9 9 9 9 9 9 9
## $ Chemical.Analysis.Method : Factor w/ 5 levels "Measured","Not coded",...: 4 4 4 4 4 4 4 4 4
## $ Chemical.Purity : Factor w/ 80 levels ">=98",">=99.0",...: 69 69 50 50 50 50 50 50
## $ Species.Scientific.Name : Factor w/ 398 levels "Acalolepta vastator",...: 69 69 248 248 248
## $ Species.Common.Name : Factor w/ 303 levels "Alfalfa Leafcutter Bee",...: 74 74 142 142
## $ Species.Group : Factor w/ 4 levels "Insects/Spiders",...: 1 1 1 1 1 1 1 1 1
## $ Organism.Lifestage : Factor w/ 20 levels "Adult","Cocoon",...: 1 1 19 19 19 1 19 1 1
## $ Organism.Age : Factor w/ 39 levels "<=24","<=48",...: 39 39 39 39 39 36 39 36 36
## $ Organism.Age.Units : Factor w/ 11 levels "Day(s)","Days post-emergence",...: 9 9 4 4 4
## $ Exposure.Type : Factor w/ 24 levels "Choice","Dermal",...: 23 23 11 11 11 11 11
## $ Media.Type : Factor w/ 10 levels "Agar","Artificial soil",...: 7 7 3 3 3 3 3
## $ Test.Location : Factor w/ 4 levels "Field artificial",...: 4 4 4 4 4 4 4 4
## $ Number.of.Doses : Factor w/ 30 levels "' 4-5',' 4-7',...: 30 30 18 18 18 18 18
## $ Conc.1.Type..Author. : Factor w/ 3 levels "Active ingredient",...: 1 1 1 1 1 1 1 1
## $ Conc.1..Author. : Factor w/ 1006 levels "<0.0004","<0.025",...: 639 510 813 622 44
## $ Conc.1.Units..Author. : Factor w/ 148 levels "%","% v/v","% w/v",...: 132 132 91 91 91 9
## $ Effect : Factor w/ 19 levels "Accumulation",...: 16 16 16 16 16 16 16
## $ Effect.Measurement : Factor w/ 155 levels "Abundance","Accuracy of learned task, per"
## $ Endpoint : Factor w/ 28 levels "EC10","EC50",...: 15 15 8 8 8 8 8 8
## $ Response.Site : Factor w/ 19 levels "Abdomen","Brain",...: 14 14 14 14 14 14 14
## $ Observed.Duration..Days. : Factor w/ 361 levels "<.0002","<.0021",...: 145 145 145 145 145
## $ Observed.Duration.Units..Days. : Factor w/ 17 levels "Day(s)","Day(s) post-emergence",...: 1 1 1
## $ Author : Factor w/ 433 levels "Abbott,V.A., J.L. Nadeau, H.A. Higo, and M"
## $ Reference.Number : int 107388 107388 103312 103312 103312 103312 103312 103312
## $ Title : Factor w/ 458 levels "A Common Pesticide Decreases Foraging Suc"
## $ Source : Factor w/ 456 levels "Acta Hortic.1094:451-456",...: 295 295 296
## $ Publication.Year : int 1982 1982 1986 1986 1986 1986 1986 1986
## $ Summary.of.Additional.Parameters: Factor w/ 943 levels "Purity: \xca NC - NC | Organism Age: \xca
```

6. Using the `summary` function on the “Effect” column, determine the most common effects that are studied. Why might these effects specifically be of interest? [Tip: The `sort()` command is useful for listing the values in order of magnitude...]

```
#What are the most common effects that are studied?
sort(summary(Neonics$Effect))
```

```
##      Hormone(s)      Histology      Physiology      Cell(s)
##           1           5           7           9
##      Biochemistry      Accumulation      Intoxication      Immunological
##           11          12          12          16
##      Morphology      Growth      Enzyme(s)      Genetics
##           22          38          62          82
##      Avoidance      Development      Reproduction      Feeding behavior
##          102          136          197          255
##      Behavior      Mortality      Population
##          360          1493          1803
```

Answer: Population is the most common effects studied. These are probably of specific interest because studying the population of an insect species can tell you a lot about the overall health and wellbeing of the species and its reaction to insecticide.

7. Using the `summary` function, determine the six most commonly studied species in the dataset (common name). What do these species have in common, and why might they be of interest over other insects? Feel free to do a brief internet search for more information if needed.[TIP: Explore the help on the `summary()` function, in particular the `maxsum` argument...]

```
#6 most commonly studied species in dataset?
sort(summary(Neonics$Species.Common.Name))
```

```
##      Ant Family      Apple Maggot
##           9           9
##      Glasshouse Potato Wasp      Lacewing
##          10          10
##      Southern House Mosquito      Two Spotted Lady Beetle
##          10          10
##      Spotless Ladybird Beetle      Braconid Parasitoid
##          11          12
##      Common Thrip      Eastern Subterranean Termite
##          12          12
##      Jassid      Mite Order
##          12          12
##      Pea Aphid      Pond Wolf Spider
##          12          12
##      Armoured Scale Family      Diamondback Moth
##          13          13
##      Eulophid Wasp      Monarch Butterfly
##          13          13
##      Predatory Bug      Yellow Fever Mosquito
##          13          13
##      Corn Earworm      Green Peach Aphid
##          14          14
```

##	House Fly	Ox Beetle
##	14	14
##	Red Scale Parasite	Spined Soldier Bug
##	14	14
##	Western Flower Thrips	Hemlock Woolly Adelgid Lady Beetle
##	15	16
##	Hemlock Woolly Adelgid	Mite
##	16	16
##	Onion Thrip	Araneoid Spider Order
##	16	17
##	Bee Order	Egg Parasitoid
##	17	17
##	Insect Class	Moth And Butterfly Order
##	17	17
##	Oystershell Scale Parasitoid	Black-spotted Lady Beetle
##	17	18
##	Calico Scale	Fairyfly Parasitoid
##	18	18
##	Lady Beetle	Minute Parasitic Wasps
##	18	18
##	Mirid Bug	Mulberry Pyralid
##	18	18
##	Silkworm	Vedalia Beetle
##	18	18
##	Codling Moth	Flatheaded Appletree Borer
##	19	20
##	Horned Oak Gall Wasp	Leaf Beetle Family
##	20	20
##	Potato Leafhopper	Tooth-necked Fungus Beetle
##	20	20
##	Argentine Ant	Beetle
##	21	21
##	Mason Bee	Mosquito
##	22	22
##	Citrus Leafminer	Ladybird Beetle
##	23	23
##	Spider/Mite Class	Tobacco Flea Beetle
##	24	24
##	Chalcid Wasp	Convergent Lady Beetle
##	25	25
##	Stingless Bee	Ground Beetle Family
##	25	27
##	Rove Beetle Family	Tobacco Aphid
##	27	27
##	Scarab Beetle	Spring Tiphia
##	29	29
##	Thrip Order	Ladybird Beetle Family
##	29	30
##	Parasitoid	Braconid Wasp
##	30	33
##	Cotton Aphid	Predatory Mite
##	33	33
##	Sweetpotato Whitefly	Aphid Family
##	37	38

##	Cabbage Looper	Buff-tailed Bumblebee
##	38	39
##	True Bug Order	Sevenspotted Lady Beetle
##	45	46
##	Beetle Order	Snout Beetle Family, Weevil
##	47	47
##	Erythrina Gall Wasp	Parasitoid Wasp
##	49	51
##	Colorado Potato Beetle	Parastic Wasp
##	57	58
##	Asian Citrus Psyllid	Minute Pirate Bug
##	60	62
##	European Dark Bee	Wireworm
##	66	69
##	Euonymus Scale	Asian Lady Beetle
##	75	76
##	Japanese Beetle	Italian Honeybee
##	94	113
##	Bumble Bee	Carniolan Honey Bee
##	140	152
##	Buff Tailed Bumblebee	Parasitic Wasp
##	183	285
##	Honey Bee	(Other)
##	667	670

```
help(summary) #what is a 'maxsum'?
sort(summary(Neonics$Species.Common.Name, maxsum = 670))
```

##	Antlike Flower Beetle	Banded Soft-winged Flower Beetle
##	1	1
##	Banded Sunflower Moth	Bee Family
##	1	1
##	Beet Armyworm	Black Citrus Aphid
##	1	1
##	Blue Alfalfa Aphid	Cabbage Root Fly
##	1	1
##	Cactus Lady Beetle	Citrus Red Mite
##	1	1
##	Cottony Cushion Sale	Crapemyrtle Aphid
##	1	1
##	Damselbug Family	Ectoparasitoid Wasp
##	1	1
##	English Grain Aphid	Fairyfly
##	1	1
##	Flea Beetle	Gall Midge
##	1	1
##	Grasshopper/Cricket/Locust Order	Greenhouse Whitefly
##	1	1
##	Grey Sunflower Seed Weevil	Harvestman Spider Order
##	1	1
##	Hawthorn Leaf Miner	Longtailed Fruit Fly Parasite
##	1	1
##	Minute Lady Beetles	Painted Maple Aphid
##	1	1

##	Pepper Weevil	Pine False Webworm
##	1	1
##	Plant Bug	Pollen Beetle
##	1	1
##	Predacious Mite	Predator Bug
##	1	1
##	Pseudocentipede Class	Pteromalid Wasp Family
##	1	1
##	Red Sunflower Seed Weevil	Rice Leaf Folder Moth
##	1	1
##	Rose Grain Aphid	Scale Picnic Beetle
##	1	1
##	Shiny Spider Beetle	Southern Army Worm
##	1	1
##	Spirea Aphid	Spotted Sunflower Stem Weevil
##	1	1
##	Strawberry Blossom Weevil	Sunflower Midge
##	1	1
##	Sunflower Moth	Ten-spot Ladybird Beetle
##	1	1
##	Tobacco Thrip	Twicestabbed Lady Beetle
##	1	1
##	Wasp Family	Weevil
##	1	1
##	Yellow Mealworm Beetle	Alfalfa Plant Bug
##	1	2
##	Alkali Bee	Aphid
##	2	2
##	Assassin Bug	Azalea Lace Bug
##	2	2
##	Banana Aphid	Brown Scale
##	2	2
##	Brown Stinkbug	Budworm
##	2	2
##	Cabbage Aphid	Cabbage White
##	2	2
##	Cardamom Thrip	Carrot Weevil
##	2	2
##	Celer Crab Spider	Centipede Class
##	2	2
##	Citricola Scale	Clouded Plant Bug
##	2	2
##	Coffee Bean Weevil	Cotton Fleahopper
##	2	2
##	Egyptian Alfalfa Weevil	Engraver Beetle
##	2	2
##	Fig Longicorn Beetle	Glassy-winged Sharpshooter
##	2	2
##	Hawthorn Lace Bug	Hister Beetle Family
##	2	2
##	Jumping Spider Family	Lined Click Beetle
##	2	2
##	Maple Spider Mite	Meshweaver Spider
##	2	2

##	Minute Pirate Bug Family	Predaceous Fly
##	2	2
##	Pygmy Mangold Beetle	Rose Sawfly
##	2	2
##	Serpentine Leafminer	Spider Mite Destroyer
##	2	2
##	Spotted Tentiform Leafminer	Stink Bug
##	2	2
##	Tawny Mole Cricket	Tick/Chigger/Mite Order
##	2	2
##	Turf Running-spider	Turnip Aphid
##	2	2
##	Western Bigeyed Bug	Western Damsel Bug
##	2	2
##	Western Plant Bug	White Apple Leafhopper Nymph
##	2	2
##	White-backed Planthopper	Whitemarked Fleahopper
##	2	2
##	Ambrosia Bark Beetle	Asian Ambrosia Beetle
##	3	3
##	Beetle Family	Birch Leafminer
##	3	3
##	Black Twig Borer	Braconid Parasitoid Wasp
##	3	3
##	California Red Scale	Crucifer Flea Beetle
##	3	3
##	Cutworm	Delphacid Planthopper
##	3	3
##	Egyptian Cotton Leafworm	Encyrtid Parasitoid
##	3	3
##	Fly/Mosquito/Midge Order	Formosan Subterranean Termite
##	3	3
##	Fruit-tree Pinhole Borer	Green Rice Leafhopper
##	3	3
##	Ground Beetle	Ichneumonid Wasp
##	3	3
##	Large-Jawed Orb Weaver Family	Leaf Cutting Ant
##	3	3
##	Mediterranean Fruit Fly	Minute Flour Bug
##	3	3
##	Mite Family	Moth Family
##	3	3
##	Negatoria Canegrub	Sap Beetle Family
##	3	3
##	Scale Insect Order	Scarab Beetle Family
##	3	3
##	Sheet-Web Weaver Family	Spider
##	3	3
##	Sugarcane Grub	Tenebrionid Beetle
##	3	3
##	Ant	Cabbage Seedpod Weevil
##	4	4
##	Common Green Lacewing	Eucalyptus Gall Wasp
##	4	4

##	European Apple Sawfly	European Honey Bee
##	4	4
##	European Tarnished Plant Bug	Garden Symphytan
##	4	4
##	Linyphiid Spider	Onion Maggot
##	4	4
##	Oriental Beetle	Parsnip Seed Wasp
##	4	4
##	Pea And Bean Weevil	Pear Sucker
##	4	4
##	Red Imported Fire Ant	Striped Cucumber Beetle
##	4	4
##	Sugarcane Beetle	Wasp
##	4	4
##	Wolf Spider Family	Yellow-faced Bumblebee
##	4	4
##	Apple Aphid	Brown Planthopper
##	5	5
##	Earwig	Green June Beetle
##	5	5
##	Hornfaced Bee	Long Horned Beetle Family
##	5	5
##	Plum Curculio	Rove Beetle
##	5	5
##	San Jose Scale	Scelionid Wasp
##	5	5
##	Speckled Cutworm Moth	Thrip Family
##	5	5
##	Ambrosia Beetle	Aphid Wasp
##	6	6
##	Black Vine Weevil	Childers Canegrub
##	6	6
##	Coconut Leaf Beetle	Eleven-spotted Ladybird Beetle
##	6	6
##	Encyrtid Wasp	European Red Mite
##	6	6
##	Fall Armyworm	Fruit Fly
##	6	6
##	Hover Fly	Oblique Banded Leaf Roller
##	6	6
##	Obscure Mealybug	Oribatid Mite Suborder
##	6	6
##	Pistachio Psyllid	Redbay Ambrosia Beetle
##	6	6
##	Silverleaf Whitefly	Soybean Aphid
##	6	6
##	Subterranean Termite	Thrip
##	6	6
##	Two-Spotted Spider Mite	Beetle Mite Family
##	6	7
##	Chinch Bug	Macedonian Honey Bee
##	7	7
##	Moth	Potato Tuberworm
##	7	7

##	Russian Wheat Aphid	Soldier Beetle
##	7	7
##	Southern One-Year Canegrub	Tarnished Plant Bug
##	7	7
##	Alfalfa Leafcutter Bee	Bee
##	8	8
##	Bumblebee	Chilean Predatory Mite
##	8	8
##	Dwarf Honey Bee	Neotropical Stingless Bee
##	8	8
##	Parasitic Wasp Family	Spiralling Whitefly
##	8	8
##	Ant Family	Apple Maggot
##	9	9
##	Asiatic Honey Bee	Eulophid Parasitoid
##	9	9
##	Lacewing Family	Mealybug Destroyer
##	9	9
##	Glasshouse Potato Wasp	Lacewing
##	10	10
##	Southern House Mosquito	Two Spotted Lady Beetle
##	10	10
##	Spotless Ladybird Beetle	Braconid Parasitoid
##	11	12
##	Common Thrip	Eastern Subterranean Termite
##	12	12
##	Jassid	Mite Order
##	12	12
##	Pea Aphid	Pond Wolf Spider
##	12	12
##	Armoured Scale Family	Diamondback Moth
##	13	13
##	Eulophid Wasp	Monarch Butterfly
##	13	13
##	Predatory Bug	Yellow Fever Mosquito
##	13	13
##	Corn Earworm	Green Peach Aphid
##	14	14
##	House Fly	Ox Beetle
##	14	14
##	Red Scale Parasite	Spined Soldier Bug
##	14	14
##	Western Flower Thrips	Hemlock Woolly Adelgid Lady Beetle
##	15	16
##	Hemlock Woolly Adelgid	Mite
##	16	16
##	Onion Thrip	Araneoid Spider Order
##	16	17
##	Bee Order	Egg Parasitoid
##	17	17
##	Insect Class	Moth And Butterfly Order
##	17	17
##	Oystershell Scale Parasitoid	Black-spotted Lady Beetle
##	17	18

##	Calico Scale	Fairyfly Parasitoid
##	18	18
##	Lady Beetle	Minute Parasitic Wasps
##	18	18
##	Mirid Bug	Mulberry Pyralid
##	18	18
##	Silkworm	Vedalia Beetle
##	18	18
##	Codling Moth	Flatheaded Appletree Borer
##	19	20
##	Horned Oak Gall Wasp	Leaf Beetle Family
##	20	20
##	Potato Leafhopper	Tooth-necked Fungus Beetle
##	20	20
##	Argentine Ant	Beetle
##	21	21
##	Mason Bee	Mosquito
##	22	22
##	Citrus Leafminer	Ladybird Beetle
##	23	23
##	Spider/Mite Class	Tobacco Flea Beetle
##	24	24
##	Chalcid Wasp	Convergent Lady Beetle
##	25	25
##	Stingless Bee	Ground Beetle Family
##	25	27
##	Rove Beetle Family	Tobacco Aphid
##	27	27
##	Scarab Beetle	Spring Tiphia
##	29	29
##	Thrip Order	Ladybird Beetle Family
##	29	30
##	Parasitoid	Braconid Wasp
##	30	33
##	Cotton Aphid	Predatory Mite
##	33	33
##	Sweetpotato Whitefly	Aphid Family
##	37	38
##	Cabbage Looper	Buff-tailed Bumblebee
##	38	39
##	True Bug Order	Sevenspotted Lady Beetle
##	45	46
##	Beetle Order	Snout Beetle Family, Weevil
##	47	47
##	Erythrina Gall Wasp	Parasitoid Wasp
##	49	51
##	Colorado Potato Beetle	Parastic Wasp
##	57	58
##	Asian Citrus Psyllid	Minute Pirate Bug
##	60	62
##	European Dark Bee	Wireworm
##	66	69
##	Euonymus Scale	Asian Lady Beetle
##	75	76

##	Japanese Beetle	Italian Honeybee
##	94	113
##	Bumble Bee	Carniolan Honey Bee
##	140	152
##	Buff Tailed Bumblebee	Parasitic Wasp
##	183	285
##	Honey Bee	
##	667	

Answer: The 6 most commonly studied species excluding the Other category are the Honey Bee, the Parasitic Wasp, the Buff Tailed Bumblebee, The Carniolan Honey Bee, the Bumble Bee, and the Italian Honeybee. All of these except for the Parasitic Wasp are in the family Apidae. These might be of interest over other insects because the family Apidae contains many important pollinators, and you wouldn't want to use insecticide on your crops and kill the insects that pollinate them. Parasitic Wasps also prey on a lot of insects that are detrimental to agricultural crops so you would not want to harm them with insecticide either.

8. Concentrations are always a numeric value. What is the class of `Conc.1..Author.` column in the dataset, and why is it not numeric? [Tip: Viewing the dataframe may be helpful...]

```
#Class of column?
str(Neonics$Conc.1..Author.)
```

```
## Factor w/ 1006 levels "<0.0004","<0.025",...: 639 510 813 622 442 637 500 642 814 784 ...
```

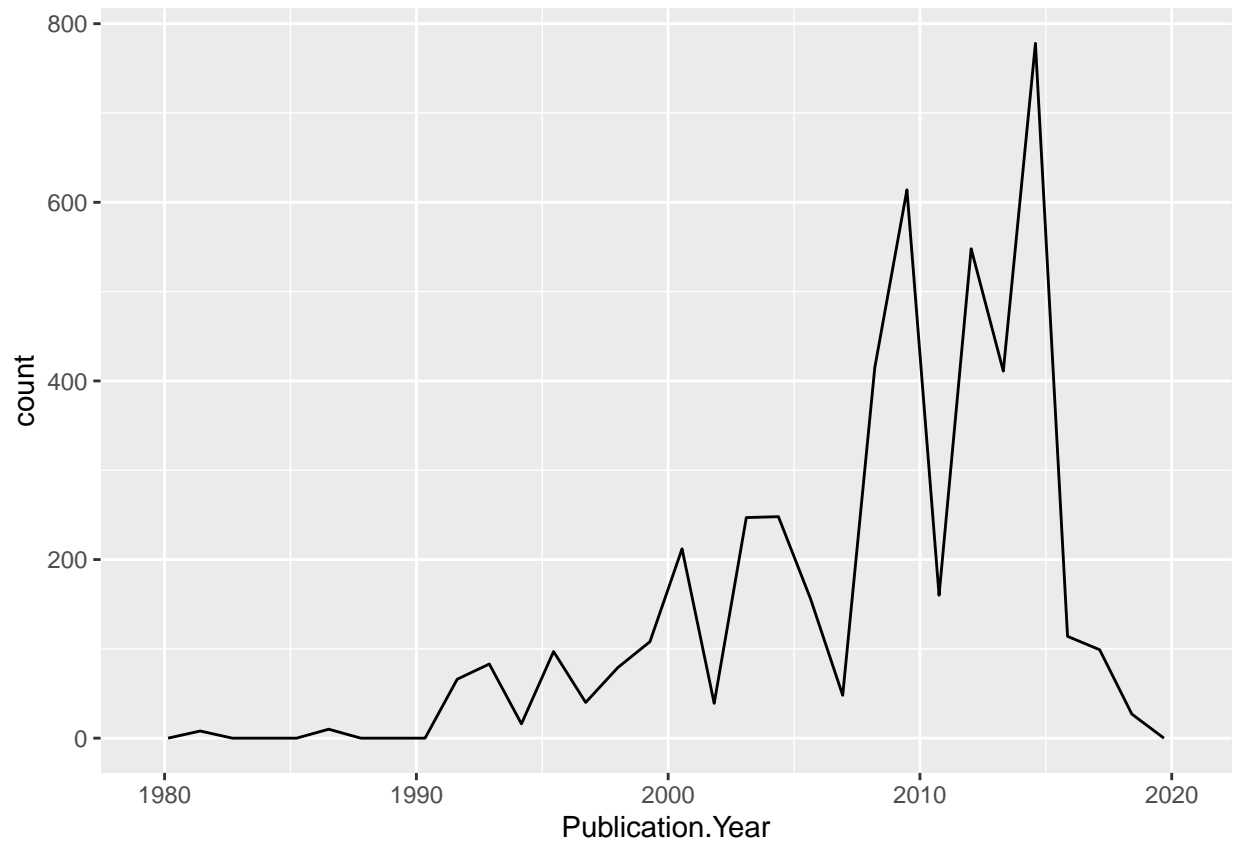
Answer: 'Conc.1..Author.' is showing up as a factor. I think this is because the numbers in this column mean different things. Some of them are the concentrations of active ingredients and some are the concentration of the formula.

Explore your data graphically (Neonics)

9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
#make plot of studies by pub. year
ggplot(Neonics) + geom_freqpoly(aes(x=Publication.Year))
```

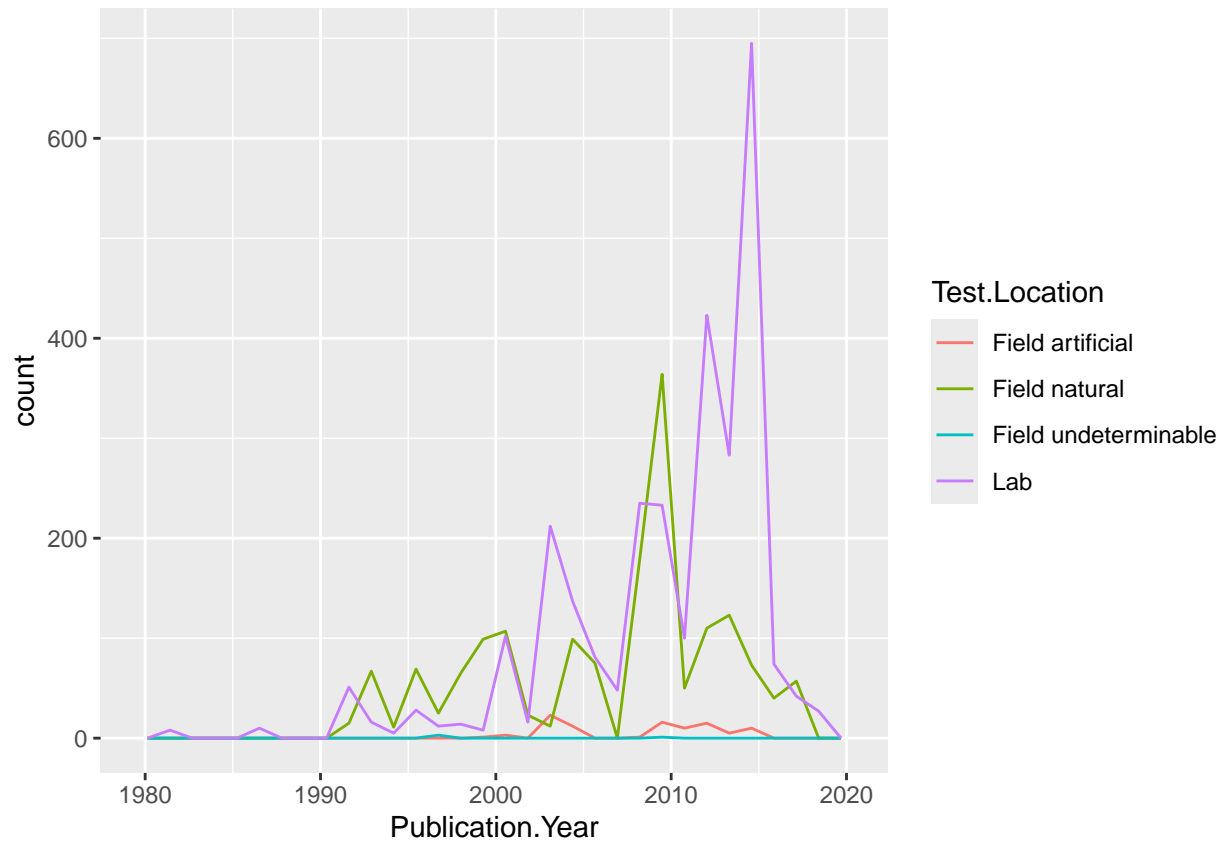
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
#Represent test locations with different colors
ggplot(Neonics) + geom_freqpoly(aes(x=Publication.Year,
                                   color=Test.Location))
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



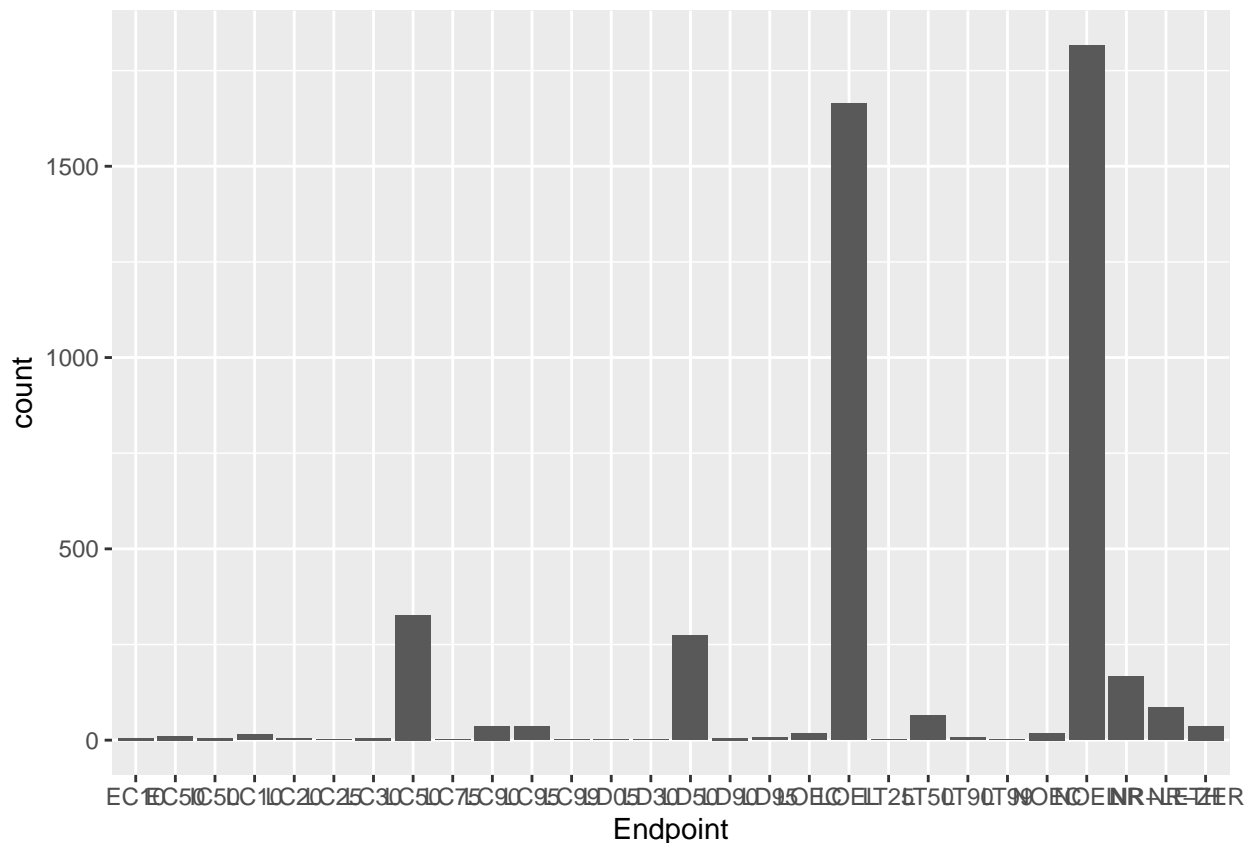
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: The most common test locations are Lab and Field natural. They do differ over time. Lab is the most common between 2010-2020 and Field Natural is most common from around 2007-2010.

11. Create a bar graph of Endpoint counts. What are the two most common end points, and how are they defined? Consult the ECOTOX_CodeAppendix for more information.

[TIP: Add `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))` to the end of your plot command to rotate and align the X-axis labels...]

```
#bar graph of Endpoint counts
ggplot(Neonics, aes(x=Endpoint)) + geom_bar();
```



```
theme(axis.text.x = element_text(
angle = 90, vjust = 0.5, hjust=1))
```

```
## List of 1
## $ axis.text.x:List of 11
## ..$ family      : NULL
## ..$ face         : NULL
## ..$ colour       : NULL
## ..$ size         : NULL
## ..$ hjust        : num 1
## ..$ vjust        : num 0.5
## ..$ angle        : num 90
## ..$ lineheight   : NULL
## ..$ margin       : NULL
## ..$ debug        : NULL
## ..$ inherit.blank: logi FALSE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## - attr(*, "class")= chr [1:2] "theme" "gg"
## - attr(*, "complete")= logi FALSE
## - attr(*, "validate")= logi TRUE
```

Answer: The two most common endpoints are LOEL and NOEL. LOEL is the lowest concentration of insecticide that had an observable effect on the insects; while, NOEL is the highest concentration of insecticide that produced no observable effect on the insects.

Explore your data (Litter)

12. Determine the class of collectDate. Is it a date? If not, change to a date and confirm the new class of the variable. Using the unique function, determine which dates litter was sampled in August 2018.

```
#What is the class of collectDate?  
str(Litter$collectDate) #collectDate is a function
```

```
## Factor w/ 2 levels "2018-08-02","2018-08-30": 1 1 1 1 1 1 1 1 1 1 ...
```

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y%m%d") #changed collectDate to a date  
unique(Litter$collectDate)
```

```
## [1] NA
```

13. Using the unique function, determine how many different plots were sampled at Niwot Ridge. How is the information obtained from unique different from that obtained from summary?

```
#using the 'unique' function vs the 'summary' function  
unique(Litter$plotID)
```

```
## [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051  
## [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057  
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

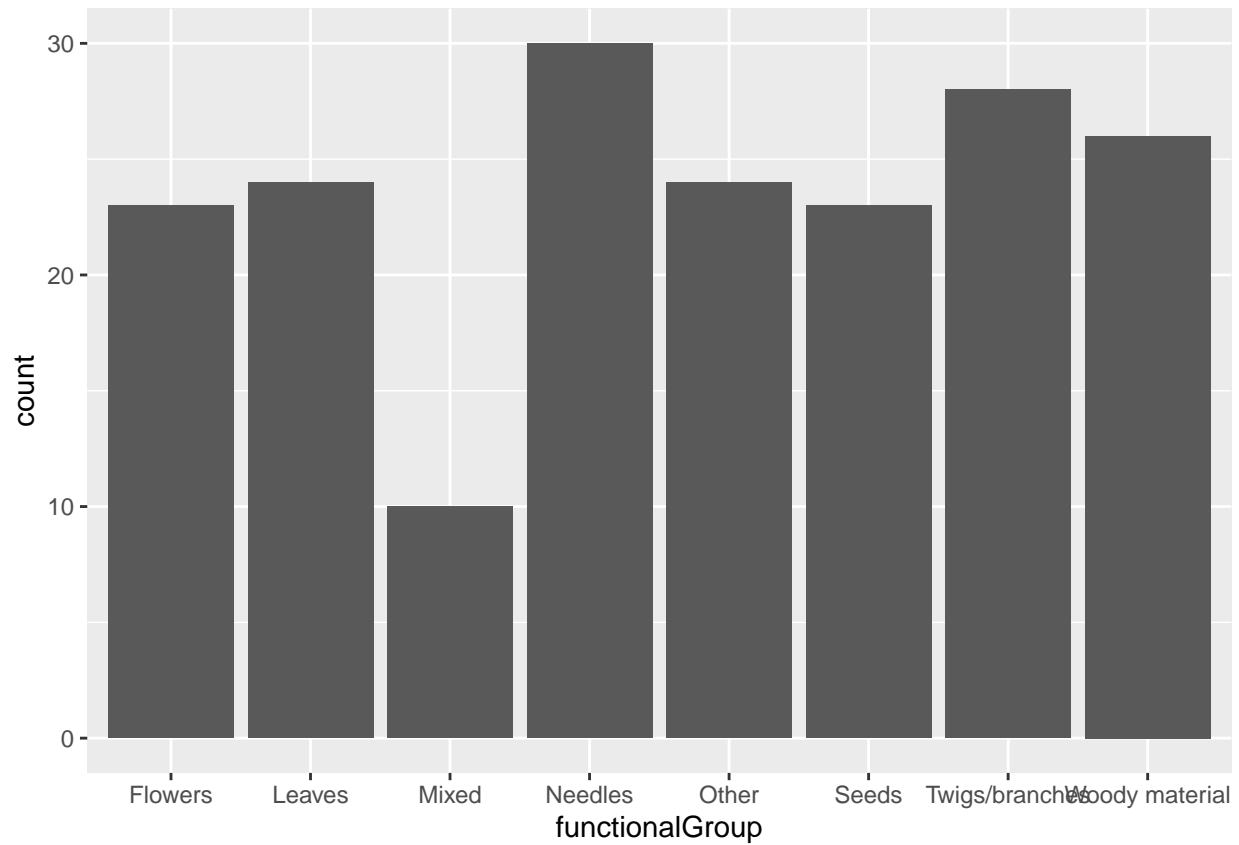
```
summary(Litter$plotID)
```

```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061  
##      20      19      18      15      14       8      16      17  
## NIWO_062 NIWO_063 NIWO_064 NIWO_067  
##      14      14      16      17
```

Answer: 12 plots were sampled at Niwot Ridge. The information obtained from ‘unique’ is the same data type as what you put into it but with the duplicate items gone. ‘summary’ shows you how many entries of each item there are in the data set.

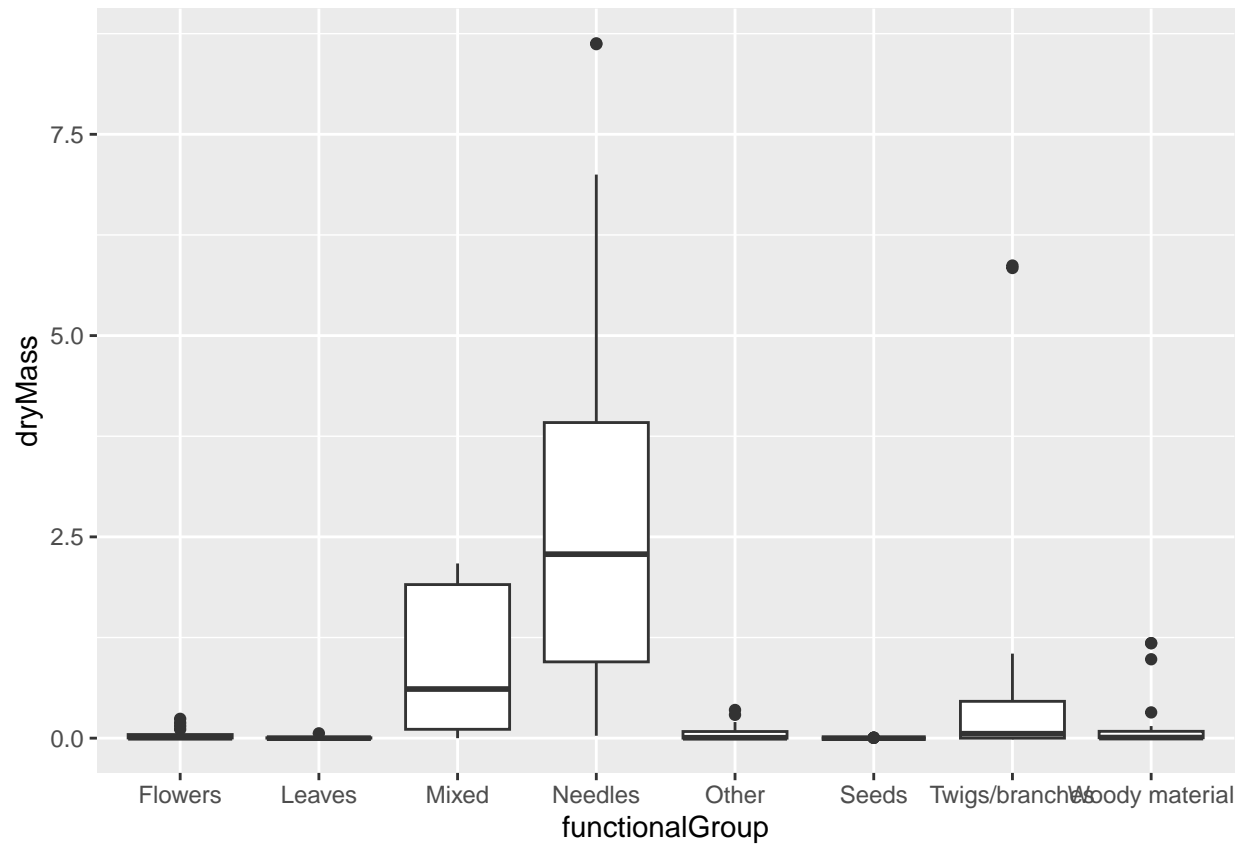
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
#bar graph of functionalGroup  
ggplot(Litter, aes(x=functionalGroup)) + geom_bar()
```

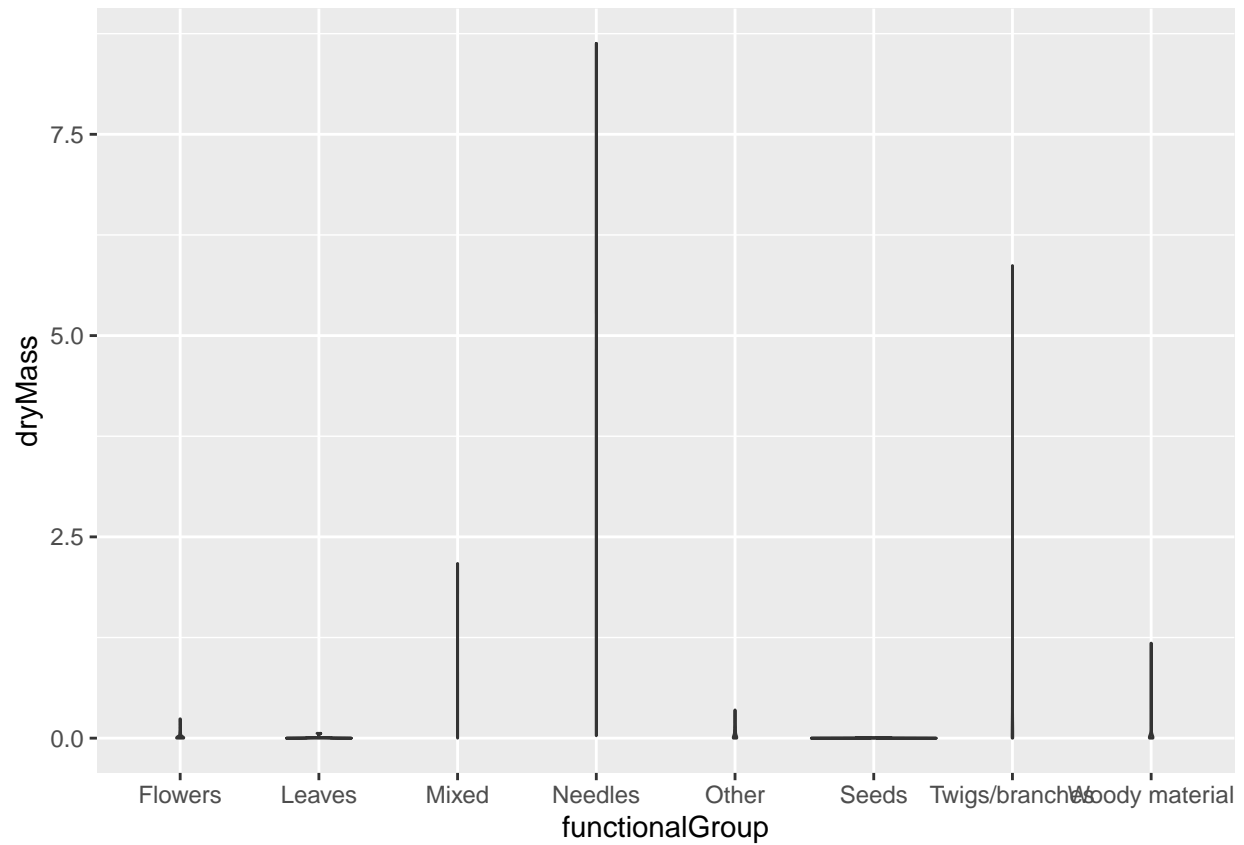



15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of `dryMass` by `functionalGroup`.

```
#boxplot and violin plot of dryMass by functionalGroup  
ggplot(Litter, aes(y=dryMass, x=functionalGroup)) + geom_boxplot()
```



```
ggplot(Litter, aes(y=dryMass, x=functionalGroup)) + geom_violin()
```



Why is the boxplot a more effective visualization option than the violin plot in this case?

Answer: This data is not the best for a violin plot because it shows up as lines. It looks much better in a box plot

What type(s) of litter tend to have the highest biomass at these sites?

Answer: Needles tend to have the highest biomass